Open Source National Dataset

Rejecting AI Companies' Free Riding on Netizens' Data Achievements

The data used by big model companies comes from the public internet, and some of it even comprises users' private data. However, the creators of this data are all internet users. It's unreasonable for these big models to utilize the creations of netizens without compensating them.

In this scenario, there are two parties involved: users and miners, with miners being numerous AI companies competing in the market. Users store their data along with corresponding wallet addresses on the blockchain, which is collectively maintained by AI enterprises to ensure its immutability. When a new entrant in the AI industry wishes to use a specific user's data (i.e., "first-time local running of a full node"), they are required to provide certain economic incentives to all users. Once miners accept and confirm the transaction, tokens are automatically distributed to the user addresses included in the block data.

So where do these newly-entered AI enterprises get ART tokens from? The answer lies in the secondary market.

Users can exercise oversight by supervising all full nodes, preventing instances of AI companies freeloading on user data. These are paid models, and users must pay tokens to use them (for example, when output reaches 10,000 tokens, they need to pay 1 ART to continue usage). Users pay the corresponding company's wallet for the model they use.

Therefore, for big model enterprises, under the condition of consistent data costs, the more powerful the model and the better the service, the more users it will attract, and the more profit the company will make.

For users, the more they participate in data creation, the more tokens they can earn. Those who contribute more data earn more, given the same frequency of usage. Additionally, rewards related to usage can be set, such as giving likes or dislikes to outputs, to promote the self-iteration of the model.

Cycle

Economic Cycle (Social Level, Promoting Comprehensive and Healthy Development of AI Iteration):

- Good user experience leads to higher profits for AI.
- Incentivized users share data more actively.
- Users earn money from data sharing.
- Models undergo self-iteration.
- Improved user experience.

Economic Cycle (Public Chain Level, Increasing the Value of the Chain):

- New model companies join.
- Purchasing tokens in the secondary market.
- Boosting the price of ART tokens.
- Users earn more.
- More active data sharing.
- More model companies join.
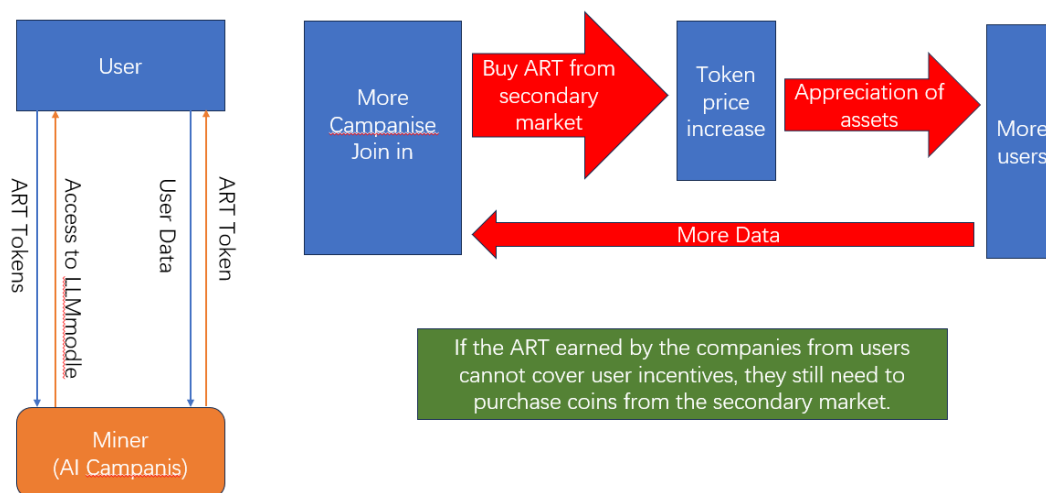
Artela's design of "elastic block space" is particularly suitable for data storage. Since user-uploaded data comes in various formats and sizes, the data size stored in each block is almost always different. Elastic blocks can adapt well to this situation.

Aspect addresses the issue of "data quality assessment."

User-uploaded data may have low quality, or even malicious intent. Therefore, determining malicious data and intercepting it, or classifying the quality of user-uploaded data to provide better incentives to users, is a critical issue.

Aspect provides highly flexible and scalable programming modules. Miners (AI model companies) can customize judgment criteria, score the data, and vote on whether to include the current data on the chain. In other words, after users upload data, it does not immediately go through packaging but first passes through the judgment module. Miners score the data based on customized judgment criteria and broadcast the scoring criteria to other miners. The mode of the scores chosen by the majority serves as the final score for the current data.

After User A broadcasts the data to all miners, the miners first use the scoring module to assign scores and then broadcast the scoring situation to other miners. Assuming a maximum score of 5 and a minimum score of 0, if the current miner scores the user's data as 2, while other miners score it as 4, then 4 will be taken as the final score for the current data and then it will be added to the blockchain.
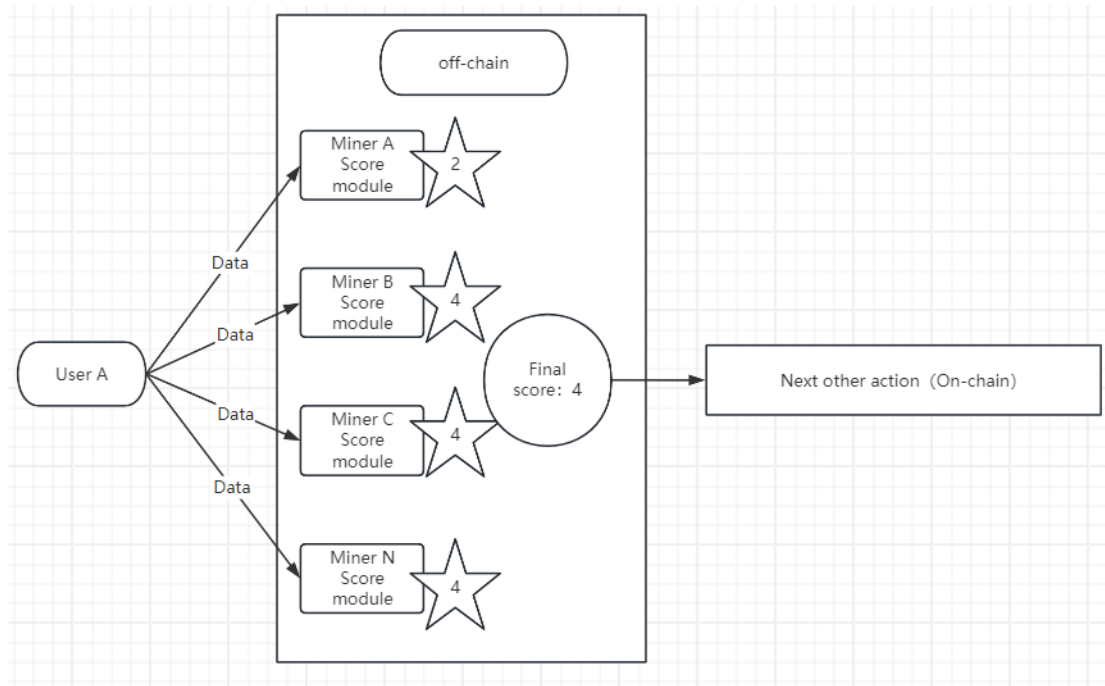
Therefore, inserting the scoring module before packaging and making the final judgment on data through voting can ensure data quality and avoid a situation where one company dominates everything. Different miners can write different judgment rules to comprehensively evaluate the quality of current data.